Reviewer#6:

Dear reviewer#6, we deeply appreciate your feedback, which has enriched our work. We now focus on your suggestions.

This is a solid paper that reports on systematic evaluation of volcanic earthquakes using several machine learning (ML) techniques. I should state at the outset that I am a volcano seismologist with more than a decade of experience, but I have never directly used any ML or AI techniques. Hence my comments are of a more general nature.

The approach in the paper is thorough. The results are repeatable, which is good. The results also show that it is possible to get more out of the data, which is always welcome. The procedures are efficient, so it is possible, in principle, to obtain similar results in much less time than it takes an experienced geophysicist to manually process the data. But this brings up a philosophical point: what is reality? Manual or ML? I would think that a manual effort by an experienced person would be the benchmark, and ML results would be judged relative to them. The paper mostly does this, with a few exceptions.

We thank the reviewer for their constructive comments and take this opportunity to clarify an important aspect of our work. The reviewer raises a philosophical question: What is reality, human or ML detection? We believe this reflection is fundamental and requires a detailed response.

From our experience, both are real and complementary. On one hand, manual effort by an experienced person serves as the benchmark but has inherent limitations, influenced by emotional and intellectual conditions due to the subjective nature of human supervision. On the other hand, the ML version is based on human benchmarks but has the advantage of conducting a systematic and rapid analysis of all information with a unified criterion, free from factors beyond what has been learned. Therefore, we can conclude that neither is more real than the other; rather, the automated method based on expert knowledge offers a competitive advantage in terms of the universality and speed of analysis, allowing for the construction of more complete catalogs but not necessarily more real ones, since those obtained manually are also real but subject to the constraints of the specific task and eruptive state in which the catalog is created.

The paper is mostly well written. Here are a few corrections keyed to line numbers:

36 – the V in VLP stand for very, not ultra

44 – is this frequency in Hz or frequency of occurrence?

51-52 – inconsistent use of parentheses ( )

143 – confront? Odd word choice

184-186 – at this point I had a hard time keeping all the acronyms in my head. I suggest adding a table of acronyms.

233 – spacing

Table 2 – are all values percentages? Needs better labeling

Table 3 – add bolder vertical lines between three main sections; spell out the abbreviations in notes at the bottom of the table (Tables should stand alone)

Table 5 – same comments as Table 3

357 – used

381 – "unbiased" but how determined? This is the place that made me rethink the question of what is reality, as described above.

Table 6 – is all the time with no events equal to the background?

In the new version of the manuscript, all these suggestions will be addressed to improve the readability and understanding of the work.

Overall, the paper is in good shape and is suitable for publication with minor revisions as indicated above.